

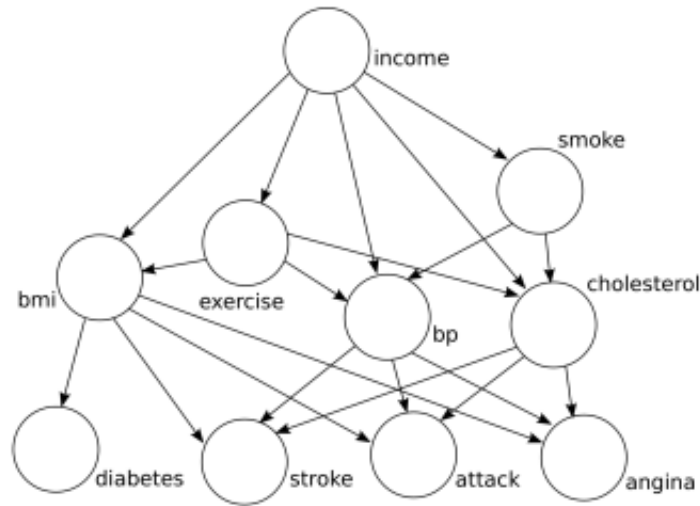
AU 332 ARTIFICIAL INTELLIGENCE: PRINCIPLES AND TECHNIQUES

By: Guowei Deng

Instructor: Yue Gao

November 18, 2018

I. INTRODUCTION



- **income** - Annual personal income level.
 1(< \$10,000) 2(\$10,000 – \$15,000) 3(\$15,000, – \$20,000)
 4(\$20,000 – \$25,000) 5(\$25,000 – \$35,000) 6(\$35,000 – \$50,000)
 7(\$50,000 – \$75,000) 8(> \$75,000)
- **exercise** - Exercised in past 30 days.
 1 (yes) 2 (no)
- **smoke** - Smoked 100 or more cigarettes in lifetime.
 1 (yes) 2 (no)
- **bmi** - Body mass index (category).
 1 (underweight) 2 (normal) 3 (overweight) 4 (obese)
- **bp** - Has high blood pressure.
 1 (yes) 2 (only when pregnant) 3 (no) 4 (pre-hypertensive)
- **cholesterol** - Has high cholesterol.
 1 (yes) 2 (no)
- **angina** - Had heart disease (angina).
 1 (yes) 2 (no)
- **stroke** - Had a stroke.
 1 (yes) 2 (no)
- **attack** - Had a heart attack.
 1 (yes) 2 (no)
- **diabetes** - Had diabetes.
 1 (yes) 2 (only during pregnancy) 3 (no) 4 (pre-diabetic)

The variables and their meanings are as above. In this lecture, we will be analyzing risk factors for certain health problems (heart disease, stroke, heart attack, diabetes) using data from the 2015 Behavioral Risk Factor Surveillance System (BRFSS) survey.

II. PROBLEMS

A. problem 1

Simply, figure out parent nodes for every node. Then create a BayesNetwork using the given function **readFactorTablefromData**. Compute the number of probabilities in every factor table. Consequently, the number of probabilities needed of this network is 504. Alternatively, the total number of probabilities needed to store the full joint distribution should be 2^{15} .

B. problem 2

The final result is as follows:

```

If I have bad habits:
The probabilities of the diabetes is
smoke exercise diabetes probs
0 1 2 1 0.150516
1 1 2 2 0.008965
2 1 2 3 0.822423
3 1 2 4 0.018096
The probabilities of the stroke is
smoke exercise stroke probs
0 1 2 1 0.049264
1 1 2 2 0.950736
The probabilities of the attack is
smoke exercise attack probs
0 1 2 1 0.07433
1 1 2 2 0.92567
The probabilities of the angina is
smoke exercise angina probs
0 1 2 1 0.080448
1 1 2 2 0.919552

```

FIG. 1: bad habits

```

If I have good habits:
The probabilities of the diabetes is
smoke exercise diabetes probs
0 2 1 1 0.127119
1 2 1 2 0.008865
2 2 1 3 0.847693
3 2 1 4 0.016323
The probabilities of the stroke is
smoke exercise stroke probs
0 2 1 1 0.03611
1 2 1 2 0.96389
The probabilities of the attack is
smoke exercise attack probs
0 2 1 1 0.052798
1 2 1 2 0.947202
The probabilities of the angina is
smoke exercise angina probs
0 2 1 1 0.054755
1 2 1 2 0.945245

```

FIG. 2: good habits

```

If I have poor health:
The probabilities of the diabetes is
bp cholesterol bmi diabetes probs
0 1 1 3 1 0.115423
1 1 1 3 2 0.007662
2 1 1 3 3 0.860873
3 1 1 3 4 0.016043
The probabilities of the stroke is
bp cholesterol bmi stroke probs
0 1 1 3 1 0.082686
1 1 1 3 2 0.917314
The probabilities of the attack is
bp cholesterol bmi attack probs
0 1 1 3 1 0.140784
1 1 1 3 2 0.859216
The probabilities of the angina is
bp cholesterol bmi angina probs
0 1 1 3 1 0.161608
1 1 1 3 2 0.838392

```

FIG. 3: poor health

```

If I have good health:
The probabilities of the diabetes is
bp cholesterol bmi diabetes probs
0 3 2 2 1 0.057710
1 3 2 2 2 0.009543
2 3 2 2 3 0.922194
3 3 2 2 4 0.010553
The probabilities of the stroke is
bp cholesterol bmi stroke probs
0 3 2 2 1 0.01446
1 3 2 2 2 0.98554
The probabilities of the attack is
bp cholesterol bmi attack probs
0 3 2 2 1 0.016161
1 3 2 2 2 0.983839
The probabilities of the angina is
bp cholesterol bmi angina probs
0 3 2 2 1 0.013326
1 3 2 2 2 0.986674

```

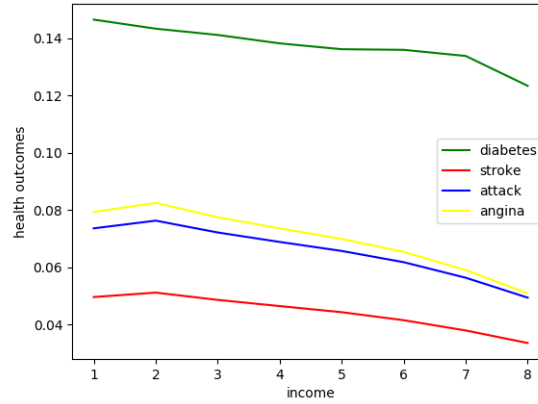
FIG. 4: good health

From the output of the code, we can learn:

health outcomes		bad habits	good habits	poor health	good health
diabetes	yes	15.05%	12.71%	11.54%	5.77%
	only during pregnancy	0.89%	0.88%	0.76%	0.95%
	no	82.24%	84.77%	86.08%	92.21%
	pre diabetic	1.81%	1.63%	1.60%	1.05%
stroke	yes	4.92%	3.61%	8.27%	1.44%
	no	95.08%	96.39%	91.73%	98.56%
heart attack	yes	7.43%	5.28%	14.08%	1.61%
	no	92.57%	94.72%	85.92%	98.39%
angina	yes	8.04%	5.47%	16.16%	1.33%
	no	91.96%	94.53%	83.84%	98.67%

C. question3

The probabilities of suffering from these 4 diseases for different income level is as follows:



Generally, we can find that the higher a person's annual income level is, the less probabilities to suffer from these diseases. However, people who earn \$10000-\$15000(level 2) rather than level 1 seem to be more likely to suffer from these 4 diseases. In terms of different diseases, the probabilities in diabetes for different income is quite close.

D. question4

If I have bad habits:

The probabilities of the diabetes is

	smoke	exercise	diabetes	is	probs
0	1	2	1	0.210945	
1	1	2	2	0.006915	
2	1	2	3	0.760693	
3	1	2	4	0.021447	

The probabilities of the stroke is

	smoke	exercise	stroke	is	probs
0	1	2	1	0.078035	
1	1	2	2	0.921965	

The probabilities of the attack is

	smoke	exercise	attack	is	probs
0	1	2	1	0.121166	
1	1	2	2	0.878834	

The probabilities of the angina is

	smoke	exercise	angina	is	probs
0	1	2	1	0.119007	
1	1	2	2	0.880993	

FIG. 5: bad habits

If I have good habits:

The probabilities of the diabetes is

	smoke	exercise	diabetes	probs
0	2	1	1	0.098552
1	2	1	2	0.009884
2	2	1	3	0.877576
3	2	1	4	0.013988

The probabilities of the stroke is

	smoke	exercise	stroke	probs
0	2	1	1	0.024311
1	2	1	2	0.975689

The probabilities of the attack is

	smoke	exercise	attack	probs
0	2	1	1	0.031015
1	2	1	2	0.968985

The probabilities of the angina is

	smoke	exercise	angina	probs
0	2	1	1	0.0368
1	2	1	2	0.9632

FIG. 6: good habits

If I have poor health:

The probabilities of the diabetes is

	bp	cholesterol	bmi	diabetes	is	probs
0	1	1	3	1	0.123481	
1	1	1	3	2	0.007460	
2	1	1	3	3	0.852416	
3	1	1	3	4	0.016643	

The probabilities of the stroke is

	bp	cholesterol	bmi	stroke	is	probs
0	1	1	3	1	0.084257	
1	1	1	3	2	0.915743	

The probabilities of the attack is

	bp	cholesterol	bmi	attack	is	probs
0	1	1	3	1	0.142199	
1	1	1	3	2	0.857801	

The probabilities of the angina is

	bp	cholesterol	bmi	angina	is	probs
0	1	1	3	1	0.162972	
1	1	1	3	2	0.837028	

FIG. 7: poor health

If I have good health:

The probabilities of the				diabetes	is
	bp	cholesterol	bmi	diabetes	probs
0	3	2	2	1	0.054173
1	3	2	2	2	0.009731
2	3	2	2	3	0.925952
3	3	2	2	4	0.010144

The probabilities of the

	bp	cholesterol	bmi	stroke	is
0	3	2	2	1	0.013997
1	3	2	2	2	0.986003

The probabilities of the

	bp	cholesterol	bmi	attack	is
0	3	2	2	1	0.015469
1	3	2	2	2	0.984531

The probabilities of the

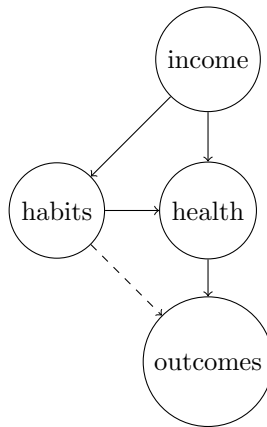
	bp	cholesterol	bmi	angina	is
0	3	2	2	1	0.012944
1	3	2	2	2	0.987056

FIG. 8: good health

TABLE I: first graph/second graph

health outcomes		bad habits	good habits	poor health	good health
diabetes	yes	15.05%/21.09%	12.71%/9.85%	11.54%/12.34%	5.77%/5.41%
	only during pregnancy	0.89%/0.69%	0.88%/0.99%	0.76%/0.75%	0.95%/0.97%
	no	82.24%/76.06%	84.77%/87.75%	86.08%/85.24%	92.21%/92.59%
	pre diabetic	1.81%/2.14%	1.63%/1.40%	1.60%/1.66%	1.05%/1.01%
stroke	yes	4.92%/7.80%	3.61%/2.43%	8.27%/8.42%	1.44%/1.40%
	no	95.08%/92.19%	96.39%/97.57%	91.73%/91.57%	98.56%/98.60%
heart attack	yes	7.43%/12.12%	5.28%/3.10%	14.08%/14.21%	1.61%/1.55%
	no	92.57%/87.88%	94.72%/96.89%	85.92%/85.78%	98.39%/98.45%
angina	yes	8.04%/11.90%	5.47%/3.68%	16.16%/16.29%	1.33%/1.29%
	no	91.96%/88.10%	94.53%/96.32%	83.84%/83.70%	98.67%/98.71%

After add edges, we can find that the probabilities in bad habits and good habits change a lot. However, the probabilities in poor health and good health in second graph is almost the same as the first graph. To clarify the relationship, we simplify the graph as follows:



Giving health's value, habits and outcomes are conditional independent in first graph. After adding edges from habits to outcomes, the probabilities of outcomes won't change a lot, which means it's not worthwhile for us to add these edges. However, given habits' values, the probabilities of outcomes change a lot, which means there are some dependences between habits and outcomes. In conclude, I think the assumption in first graph is not valid. Without the edges from habits to outcomes, the graph is not accurate.

E. question5

Question5:			
diabetes	stroke	probs	
0	1	1	0.044164
1	1	2	0.955836
diabetes	stroke	probs	
0	3	1	0.040478
1	3	2	0.959522
After adding edge from diabetes to stroke			
diabetes	stroke	probs	
0	1	1	0.076198
1	1	2	0.923802
diabetes	stroke	probs	
0	3	1	0.035015
1	3	2	0.964985

From the result below, we find that if a person has diabetes, it's much more likely that he or she has a stroke(7.61% vs 4.41%). So I think the interaction between diabetes and stroke are valid.